Biometrika (2017), **104**, 2, *pp.* 439–453
Printed in Great Britain
doi: 10.1093/biomet/asx007
Advance Access publication 17 March 2017

Multiply robust imputation procedures for the treatment of item nonresponse in surveys

By SIXIA CHEN

Department of Biostatistics and Epidemiology, University of Oklahoma, 801 NE 13th Street, Oklahoma City, Oklahoma 73104, U.S.A.

sixia-chen@ouhsc.edu

and DAVID HAZIZA

Department of Mathematics and Statistics, Université de Montréal, Pavillon André-Aisenstadt, 2920 chemin de la Tour, Montreal H3C 3J7, Canada

david.haziza@umontreal.ca

Summary

Item nonresponse in surveys is often treated through some form of imputation. We introduce multiply robust imputation in finite population sampling. This is closely related to multiple robustness, which extends double robustness. In practice, multiple nonresponse models and multiple imputation models may be fitted, each involving different subsets of covariates and possibly different link functions. An imputation procedure is said to be multiply robust if the resulting estimator is consistent when all models but one are misspecified. A jackknife variance estimator is proposed and shown to be consistent. Random and fractional imputation procedures are discussed. A simulation study suggests that the proposed estimation procedures have low bias and high efficiency.

Some key words: Double robustness; Imputation; Item nonresponse; Jackknife; Model calibration; Survey data.

1. Introduction

In surveys conducted by statistical agencies, item nonresponse is usually handled through some form of imputation. Most often, a single imputation procedure is used, whereby a missing value is replaced by a single imputed value, which is constructed using auxiliary information in the form of a set of fully observed variables. Imputation is used to reduce the nonresponse bias and to produce a complete rectangular data file, allowing secondary analysts to obtain point estimates using complete-data estimation procedures. Any estimation procedure that aims at reducing the nonresponse bias requires assumptions. Doubly robust procedures have attracted a lot of attention in recent years; e.g., Robins et al. (1994), Scharfstein et al. (1999), Tan (2006), Bang & Robins (2005), Kang & Schafer (2007), and Cao et al. (2009). In the context of survey data, doubly robust procedures have been discussed in Kott (1994), Kott (2006), Kim & Park (2006), Haziza & Rao (2006), Kott & Chang (2010), Haziza et al. (2014), and Kim & Haziza (2014), among others. An estimator is said to be doubly robust if it remains asymptotically unbiased if either the nonresponse model or the imputation model is correctly specified. A nonresponse model is a set of assumptions about the unknown nonresponse mechanism, whereas an imputation model, also called an outcome regression model, is a set of assumptions about the distribution of the study variable being imputed. Doubly robust procedures offer some protection if either model is misspecified, but as the missing data-generating process is unknown, they may not provide sufficient protection because it is risky to assume that either model is correctly specified (Han, 2014b).

To cope with this problem, Han & Wang (2013) introduced the concept of multiple robustness in an infinite population set-up; see also Han (2014a,b) and Chan & Yam (2014). Multiple robustness can be viewed as an extension of double robustness. In practice, multiple nonresponse models and multiple imputation models may be fitted, each involving different subsets of covariates and possibly different link functions. An estimation procedure is said to be multiply robust when it is consistent if any one of those multiple models, for either the response probability or the study variable, is correctly specified. This is attractive because it provides protection against model misspecification.

Multiply robust procedures may be useful in a number of situations. Assume that we are interested in estimating the finite population total of a study variable y subject to missingness and let x be a vector of fully observed variables. It may be difficult to build a model describing the relationship between y and x that places all other models out of consideration. Even using recent techniques of variable selection, different levels of tuning may lead to different models. Building an enlarged model would be an option, but estimation may be problematic if the dimension of x is very large. In practice, it seems desirable to select a set of reasonable models, and to incorporate them simultaneously. In the context of survey data, it is important to account for characteristics of the survey design, such as stratification, sampling weights and clustering. With multiple models, one does not have to decide whether or not to include the sampling weights in the model; they may be included in some models and excluded from others. Perhaps none of the selected models is correctly specified, but, as pointed out by Han (2014b), multiply robust procedures, even when inconsistent, tend to have good numerical performance even if all the models are misspecified. This is usually not the case if the estimation procedure is based on a single model; see § 7. The idea of fitting multiple models in the context of a large number of predictors was first suggested in Robins et al. (2007).

Suppose that the variable y to be imputed is binary, with $y = 1$ if the sample unit has a characteristic of interest and $y = 0$ otherwise. In practice, it is customary to replace the missing value by simulating from a fitted logistic regression model, but if the link function is misspecified, the resulting estimator may be biased. This can be overcome by using a nonparametric procedure such as kernel or local polynomial regression (Wand & Jones, 1995), but unless the number of predictors is very small, one may face the curse of dimensionality. In such cases, multiply robust procedures are attractive because one can fit multiple models, each corresponding to a specific link function.

Transforming response and/or predictor variables prior to imputation may remedy problems such as asymmetry, nonlinearity and unequal variances; see, for example, Fox (2008). Transformation in the context of imputation is discussed in He & Raghunathan (2006) and Robbins et al. (2013), among others. The latter authors examine the effect of data transformation in the context of the Agricultural Resource Management Survey; see Weisberg (1985) for a discussion about transformations of the response or the predictors. Choosing one transformation may be problematic unless the sample is very large (Léger et al., 1992), so rather than using a single transformation, it seems attractive to fit multiple models, each corresponding to a particular transformation. This is illustrated empirically in § 7.

2. Theoretical set-up

Consider a finite population U of size N. We focus on estimating the population total $Y = \sum_{i\in U} y_i$ of a study variable y. We select a sample s of size n according to a sampling design $p(s)$. In the absence of nonresponse, a complete-data estimator of Y is the expansion estimator $\hat{Y}_\pi = \sum_{i\in s} w_i y_i$, where $w_i = 1/\pi_i$ denotes the design weight attached to unit i and π_i denotes its inclusion probability in the sample. We assume that $\pi_i > 0$ for all units. The expansion estimator is design-unbiased and design-consistent for Y (e.g., Isaki & Fuller, 1982).

In the presence of nonresponse to the study variable y, an estimator of Y is defined as

$$\hat{Y}_{\rm I} = \sum_{i\in s} w_i r_i y_i + \sum_{i\in s} w_i(1-r_i)y_i^*, \tag{1}$$

where y_i^* denotes the imputed value used to replace the missing value y_i and r_i is a response indicator such that $r_i = 1$ if unit i is a respondent to the study variable y and $r_i = 0$ otherwise. We assume that the first component of the vector of fully observed variables, x, is equal to 1 for all the sample units. Throughout the paper, we assume that the missing-at-random assumption (Rubin, 1976) holds:

$$\text{pr}(r_i = 1 \mid x_i, y_i) = \text{pr}(r_i = 1 \mid x_i) \equiv p_i.$$

In order to construct the imputed values y_i^*, we postulate the imputation model

$$y_i = m(x_i;\beta) + \epsilon_i, \tag{2}$$

where β is a vector of unknown coefficients. We assume that $E_m(\epsilon_i) = 0$, $E_m(\epsilon_i\epsilon_j) = 0\ (i \neq j)$ and $\text{var}(\epsilon_i) = \sigma^2$, where the subscript m refers to the imputation model (2). Although we assume equal variances, our results extend to unequal variances.

Let s_r denote the set of respondents to the study variable y, of size n_r, and let $s_m = s - s_r$ denote the set of nonrespondents, of size n_m, such that $s = s_r \cup s_m$ and $n = n_r + n_m$. Deterministic imputation consists of replacing the missing value y_i by $y_i^* = m(x_i;\hat\beta)$ $(i \in s_m)$, where $\hat\beta$ is a solution of the estimating equations

$$\sum_{i\in s} \phi_i r_i \left\{y_i - m\left(x_i;\beta\right)\right\} \frac{\partial m(x_i;\beta)}{\partial\beta} = 0 \tag{3}$$

and ϕ_i in (3) is a coefficient attached to unit i. Regardless of the choice of ϕ_i, the imputed estimator (1) is consistent for Y if the imputation model (2) holds. The choice $\phi_i = w_i$ leads to survey-weighted deterministic imputation, whereas the choice $\phi_i = 1$ leads to unweighted deterministic imputation. However, with these choices, the imputed estimator (1) is generally inconsistent if the imputation model (2) is misspecified. To cope with this problem, an alternative choice of ϕ_i can be obtained by postulating a nonresponse model,

$$p_i = p(x_i;\alpha), \tag{4}$$

where α is a vector of unknown coefficients. Let $\hat p_i = p(x_i;\hat\alpha)$ denote the estimated response probability for unit i, where $\hat\alpha$ is a solution of the estimating equations

$$\sum_{i\in s} w_i \frac{r_i - p(x_i;\alpha)}{p(x_i;\alpha)\left\{1 - p(x_i;\alpha)\right\}} \frac{\partial p(x_i;\alpha)}{\partial\alpha} = 0.$$

The choice $\phi_i = w_i(\hat{p}_i^{-1} - 1)$ ensures that the imputed estimator (1) is consistent for Y provided that the nonresponse model (4) is correctly specified, regardless of whether or not the imputation model (2) is correctly specified. Hence, with the choice $\phi_i = w_i(\hat{p}_i^{-1} - 1)$, the resulting imputed estimator, denoted by $\hat{Y}_{\rm DR}$, is doubly robust; see, for example, Haziza & Rao (2006). Rather than fitting a single nonresponse model and/or a single imputation model, it may be desirable to fit multiple nonresponse models and multiple imputation models.

3. Proposed method

Let $\mathcal{C}_1 = \{p^j(x_i;\alpha^j) : j = 1,\ldots,J\}$ be a set of J nonresponse models and $\mathcal{C}_2 = \{m^k(x_i;\beta^k) : k = 1,\ldots,K\}$ be a set of K imputation models. The corresponding estimators $\hat{\alpha}^j$ and $\hat{\beta}^k$ are obtained by solving the survey-weighted estimating equations

$$S_1^j(\alpha^j) = \sum_{i\in s} w_i \frac{r_i - p^j(x_i;\alpha^j)}{p^j(x_i;\alpha^j)\left\{1 - p^j(x_i;\alpha^j)\right\}} \frac{\partial p^j(x_i;\alpha^j)}{\partial \alpha^j} = 0,$$

$$S_2^k(\beta^k) = \sum_{i\in s} w_i r_i \left\{y_i - m^k(x_i;\beta^k)\right\} \frac{\partial m^k(x_i;\beta^k)}{\partial \beta} = 0.$$

Our imputation procedure has two steps. In the first step, we obtain calibrated weights $\tilde{w}_i$ as close as possible to the initial weights w_i such that the following $J + K + 1$ calibration constraints are satisfied:

$$\sum_{i\in s_r} \tilde{w}_i = \sum_{i\in s} w_i, \tag{5}$$

$$\sum_{i\in s_r} \tilde{w}_i L\left\{1/p^j(x_i;\hat{\alpha}^j)\right\} = \sum_{i\in s} w_i L\left\{1/p^j(x_i;\hat{\alpha}^j)\right\}, \tag{6}$$

$$\sum_{i\in s_r} \tilde{w}_i m^k(x_i;\hat{\beta}^k) = \sum_{i\in s} w_i m^k(x_i;\hat{\beta}^k), \tag{7}$$

where $L(t)$ is the inverse function of $F(t)$, which is a calibration function defined below in (8). The calibration constraints (5)–(7) are similar to those encountered in the context of model calibration for complete data (Wu & Sitter, 2001). More specifically, we seek calibrated weights $\tilde{w}_i$ such that

$$\sum_{i\in s_r} G(\tilde{w}_i/w_i)$$

is minimized subject to (5)–(7), where $G(\tilde{w}_i/w_i)$ is a strictly convex function, differentiable with respect to $\tilde{w}_i$ with continuous derivatives $g(\tilde{w}_i/w_i) = \partial G(\tilde{w}_i/w_i)/\partial \tilde{w}_i$ and such that $g(1) = 0$. Furthermore, the function satisfies $G(\tilde{w}_i/w_i) \geqslant 0$ and $G(1) = 0$; see Deville & Särndal (1992). Popular distance functions include the generalized chi-square distance $G(\tilde{w}_i/w_i) = (1/2)(\tilde{w}_i/w_i - 1)^2$, the generalized pseudo-empirical likelihood distance $G(\tilde{w}_i/w_i) = -\log(\tilde{w}_i/w_i) + \tilde{w}_i/w_i - 1$, and the generalized exponential tilting distance $G(\tilde{w}_i/w_i) = (\tilde{w}_i/w_i)\log(\tilde{w}_i/w_i) - \tilde{w}_i/w_i + 1$; see Wu & Lu (2016).

The weights $\tilde{w}_i$ are given by

$$\tilde{w}_i = w_i F(\hat{\lambda}_r^{\rm T} h_i), \quad i \in s_r, \tag{8}$$

where $F(\cdot)$ is a calibration function, defined as the inverse function of $g(\cdot)$, $\hat{\lambda}_r$ is a $(J+K+1)$-vector of estimated coefficients satisfying (5)–(7), and

$$h_i = \left(1, \hat{L}_i^1 - \hat{L}^1, \ldots, \hat{L}_i^J - \hat{L}^J, \hat{m}_i^1 - \hat{m}^1, \ldots, \hat{m}_i^K - \hat{m}^K\right)^{\mathrm{T}} \tag{9}$$

with $\hat{L}_i^j \equiv L\left\{1/p^j(x_i;\hat{\alpha}^j)\right\}$, $\hat{m}_i^k \equiv m^k(x_i;\hat{\beta}^k)$ and

$$\hat{L}^j \equiv \frac{\sum_{i\in s} w_i L\left\{1/p^j(x_i;\hat{\alpha}^j)\right\}}{\sum_{i\in s} w_i}, \quad \hat{m}^k \equiv \frac{\sum_{i\in s} w_i m^k(x_i;\hat{\beta}^k)}{\sum_{i\in s} w_i} \quad (j=1,\ldots,J;\ k=1,\ldots,K).$$

The generalized chi-square distance leads to $\hat{L}_i^j = 1/p^j(x_i;\hat{\alpha}^j)$ and (8) reduces to $\tilde{w}_i = w_i(1+\hat{\lambda}_r^{\mathrm{T}}h_i)$. The generalized pseudo-empirical likelihood distance leads to $\hat{L}_i^j = p^j(x_i;\hat{\alpha}^j)$ and $\tilde{w}_i = w_i(1+\hat{\lambda}_r^{\mathrm{T}}h_i)^{-1}$. Finally, the generalized exponential tilting distance leads to $\hat{L}_i^j = \log\left\{p^j(x_i;\hat{\alpha}^j)\right\}$ and $\tilde{w}_i = w_i\exp(\hat{\lambda}_r^{\mathrm{T}}h_i)$. With the generalized chi-square distance, some weights $\tilde{w}_i$ may be negative. Both the generalized pseudo-empirical likelihood distance and the generalized exponential tilting distance ensure that $\tilde{w}_i > 0$ for all i, although some weights may be extreme. The problem of extreme weights can be overcome by using a distance function that ensures that the weights lie between prespecified lower and upper bounds; see Deville & Särndal (1992). For brevity, we write $\hat{F}_i$ for $F(\hat{\lambda}_r^{\mathrm{T}}h_i)$ in the following.

In the second step, the imputed values y_i^* are obtained by fitting a weighted linear regression with y as the dependent variable and h_i given by (9) as the vector of independent variables. The regression weights are given by $w_i(\hat{F}_i - 1)$. This leads to

$$y_i^* = h_i^{\mathrm{T}}\hat{\gamma}_p, \quad i \in s_m, \tag{10}$$

where

$$\hat{\gamma}_p = \left\{\sum_{i\in s} r_i w_i\left(\hat{F}_i - 1\right)h_i h_i^{\mathrm{T}}\right\}^{-1}\left\{\sum_{i\in s} r_i w_i\left(\hat{F}_i - 1\right)h_i y_i\right\}. \tag{11}$$

The resulting imputed estimator,

$$\hat{Y}_{\mathrm{MR}} = \sum_{i\in s} r_i w_i y_i + \sum_{i\in s}(1-r_i)\, w_i h_i^{\mathrm{T}}\hat{\gamma}_p, \tag{12}$$

can be implemented from the imputed data file containing the survey weights w_i and the values $\tilde{y}_i = r_i y_i + (1-r_i)h_i^{\mathrm{T}}\hat{\gamma}_p$. Because the first component of the vector h_i is equal to 1 for all i, (12) equals

$$\hat{Y}_{\mathrm{MR}} = \sum_{i\in s_r} w_i\hat{F}_i y_i + \left(\sum_{i\in s} w_i h_i - \sum_{i\in s_r} w_i\hat{F}_i h_i\right)^{\mathrm{T}}\hat{\gamma}_p = \sum_{i\in s_r} w_i\hat{F}_i y_i. \tag{13}$$

The second equality in (13) follows from the calibration constraints (5)–(7). In the next section, we show that $\hat{Y}_{\mathrm{MR}}$ is multiply robust. For this reason, (10) is referred to as a multiply robust deterministic imputation procedure.

4. Asymptotic results

The following theorem establishes the consistency of $\hat{Y}_{\mathrm{MR}}$ when one of the imputation models is true or when the true model is a linear combination of the multiple imputation models.

Theorem 1. *If one of the imputation models is true or the true model is a linear combination of the multiple imputation models, then $\hat{Y}_{\mathrm{MR}}/Y$ converges in probability to 1 as n and N go to infinity.*

A sketch of the proof is presented in the Appendix. The following theorem establishes the consistency of $\hat{Y}_{\mathrm{MR}}$ when one of the nonresponse models is true.

Theorem 2. *If one of the nonresponse models is true, then $\hat{Y}_{\mathrm{MR}}/Y$ converges in probability to 1 as n and N go to infinity.*

A sketch of the proof is presented in the Appendix. Combining Theorems 1 and 2, we conclude that $\hat{Y}_{\mathrm{MR}}$ is multiply robust, in the sense that it is consistent if all but one of the models are misspecified. The next theorem presents an asymptotic expression for $\hat{Y}_{\mathrm{MR}}$.

Theorem 3. *Under the regularity conditions* A1–A4 *in the Appendix,*

$$\hat{Y}_{\mathrm{MR}} = \sum_{i\in s} w_i \eta_{i,0} + O_{\mathrm{p}}(Nn^{-1}),$$

where

$$\eta_{i,0} = r_i F_i y_i + A_1(1 - r_i F_i) h_{i,0} + A_2 Q(x_i;\theta^*),$$

with

$$A_1 = 1 + E\left(\sum_{i\in s} w_i r_i \dot{F}_i h_{i,0} e_{i,0}\right)\left\{E\left(\sum_{i\in s} w_i r_i \dot{F}_i h_{i,0} h_{i,0}^{\mathrm{T}}\right)\right\}^{-1},$$

$$A_2 = E\left\{\frac{1}{N}\left(\sum_{i\in s} w_i r_i \dot{F}_i \lambda^* \dot{h}_{i,0} e_{i,0} - \sum_{i\in s} w_i (r_i F_i - 1) r_i \dot{h}_{i,0}\right)\right\}.$$

Here the expectations are evaluated with respect to the sampling design, the nonresponse mechanism and the true imputation model, $Q(x_i;\theta^*)$ is defined in the Appendix, $h_{i,0} = h(x_i;\theta^*)$, $\dot{h}_{i,0} = \dot{h}(x_i;\theta^*)$, $\dot{h}(x_i;\theta^*) = \partial h(x_i;\theta)/\partial\theta$ evaluated at $\theta = \theta^*$, $e_{i,0} = y_i - \gamma^{*\mathrm{T}} h_{i,0}$ and $\dot{F}_i = \partial F(t)/\partial t$ evaluated at $t = \lambda^* h_{i,0}$.

A sketch of the proof is presented in the Appendix.

Remark 1. In the special case of a single nonresponse model and a single imputation model, the estimator (12) is different from the doubly robust imputation estimator discussed in Haziza & Rao (2006). This is illustrated empirically in § 7.

5. Variance estimation

When the sampling fraction n/N is negligible, the variance of $\hat{Y}_{\rm MR}$ can be estimated using the generalized jackknife of Berger (2007). In the complete-data case, Berger (2007) proposed a jackknife procedure for unequal probability sampling designs and showed that it is design-consistent under the conditional Poisson sampling design, which is the maximum entropy design. Berger (2011) showed that if an estimator is consistent under the conditional Poisson sampling design, then it is consistent under any sampling design close to the maximum entropy design. Examples of sampling designs close to conditional Poisson sampling include the Rao–Sampford design (Rao, 1965; Sampford, 1967) and randomized systematic sampling (Tillé, 2006, § 7.2).

Let $w_{i(j)}$ be the jackknife weights, with $w_{i(j)} = n(n-1)^{-1}w_i$ if $i \neq j$ and $w_{i(j)} = 0$ if $i = j$. Let $\hat{Y}_{{\rm MR}(j)}$ denote the estimator $\hat{Y}_{\rm MR}$ based on the data with the jth unit excluded,

$$\hat{Y}_{{\rm MR}(j)} = \sum_{i\in s} w_{i(j)} r_i y_i + \sum_{i\in s} w_{i(j)}(1-r_i) h_{i(j)}^{\rm T} \hat{\gamma}_{p(j)},$$

where $\hat{\gamma}_{p(j)}$ is computed in the same way as $\hat{\gamma}_p$ in (11) but with the jackknife weights $w_{i(j)}$ instead of the original weights w_i and with h_i replaced by $h_{i(j)}$. Computing $h_{i(j)}$ involves fitting the J nonresponse models and the K imputation models based on the data with the jth unit excluded. A jackknife variance estimator is

$$\hat{V}_{\rm J} = \frac{n}{n-1}\sum_{i\in s}(1-\pi_i)\left(u_i - \sum_{k\in s}\varphi_k u_k\right)^2, \tag{14}$$

where $u_i = (1-w_i)(\hat{Y}_{\rm MR} - \hat{Y}_{{\rm MR}(i)})$ and $\varphi_i = c_i/\sum_{k\in s} c_k$ with $c_i = n/(n-1)(1-\pi_i)$.

Using the results of Berger (2007, 2011), the jackknife variance estimator (14) is consistent for the true variance provided that the sampling fraction n/N is negligible. The consistency of $\hat{V}_{\rm J}$ can be established through the variance estimation approach of Fay (1991); see also Shao & Steel (1999), Haziza (2009) and Kim & Rao (2009). This result holds for the conditional Poisson sampling design and any other sampling design close to the maximum entropy design in the sense of Berger (2011). Also, $\hat{V}_{\rm J}$ is consistent even if all the nonresponse models and all the imputation models are misspecified. That is, the jackknife variance estimator (14) tracks the true variance of $\hat{Y}_{\rm MR}$, be the latter biased or unbiased. A $(1-\alpha\%)$ confidence interval for Y is

$$\hat{Y}_{\rm MR} \pm z_{\alpha/2}\hat{V}_{\rm J}^{1/2}, \tag{15}$$

where $z_{\alpha/2}$ denotes the upper $(1-\alpha/2)$ critical value for the standard normal distribution. The confidence interval (15) is multiply robust in the sense that its coverage probability is close to nominal if all but one of the models are misspecified.

6. Multiply robust random and fractional imputation

6·1. *Random imputation*

We consider a random imputation procedure that consists of replacing the missing y_i by

$$y_{ij}^* = h_i^{\rm T}\hat{\gamma}_p + e_j^*, \tag{16}$$

where e_j^* is selected at random from the residuals observed among the responding units, $\{y_j - h_j^{\mathrm{T}}\hat{\gamma}_p,\ j \in s_r\}$, with probability

$$w_{ij}^* = \frac{w_j(\hat{F}_j - 1)}{\sum_{k\in s} r_k w_k(\hat{F}_k - 1)}.$$

The resulting imputed estimator obtained by using (16) in (1) is denoted by $\hat{Y}_{\mathrm{MRR}}$. The latter is multiply robust as $E(e_j^*) = 0$. Its variance can be expressed as

$$\mathrm{var}(\hat{Y}_{\mathrm{MRR}}) = \mathrm{var}(\hat{Y}_{\mathrm{MR}}) + E\left\{\mathrm{var}\left(\hat{Y}_{\mathrm{MRR}} - \hat{Y}_{\mathrm{MR}}\right)\right\}, \tag{17}$$

where the first term on the right-hand side of (17) denotes the variance of $\hat{Y}_{\mathrm{MR}}$ under (10), and the second term denotes the imputation variance arising from the random selection of the residuals e_j^*. A multiply robust variance estimator of $\mathrm{var}(\hat{Y}_{\mathrm{MRR}})$ is given by

$$\widehat{\mathrm{var}}(\hat{Y}_{\mathrm{MRR}}) = \widehat{\mathrm{var}}_{\mathrm{J}} + \widehat{\mathrm{var}}_{\mathrm{I}},$$

where $\widehat{\mathrm{var}}_{\mathrm{J}}$ is given by (14) and

$$\widehat{\mathrm{var}}_{\mathrm{I}} = \sum_{i\in s} w_i^2(1 - r_i)\sum_{j\in s_r} w_{ij}^*\left(y_{ij}^* - \sum_{j\in s_r} w_{ij}^* y_{ij}^*\right)^2.$$

6·2. *Fractional imputation*

We now turn to fractional imputation, as considered in Kim & Fuller (2004) and Fuller & Kim (2005). The imputed estimator (1) under fractional imputation can be written as

$$\hat{Y}_{\mathrm{MRF}} = \sum_{i\in s} w_i r_i y_i + \sum_{i\in s} w_i(1 - r_i)\sum_{j\in s_r} w_{ij}^* y_{ij}^*, \tag{18}$$

where y_{ij}^* is given by (16) and w_{ij}^* denotes the fraction of the original weight of recipient i assigned to the value from donor j. We have $\sum_{j\in s_r} w_{ij}^* = 1$, $w_{ij}^* \geqslant 0$. The estimator (18) can be rewritten as

$$\hat{Y}_{\mathrm{MRF}} = \sum_{i\in s} w_i h_i^{\mathrm{T}}\hat{\gamma}_p + \sum_{i\in s} r_i\left\{w_i + \sum_{j\in s}(1 - r_j)w_j w_{ij}^*\right\} e_i^*. \tag{19}$$

Comparing (19) with (13), we have

$$w_i + \sum_{j\in s}(1 - r_j)w_j w_{ij}^* = w_i\hat{F}_i. \tag{20}$$

Because $\sum_{j\in s}(1 - r_j)w_j = \sum_{j\in s} r_j w_j(\hat{F}_j - 1)$, the fractional weights satisfying (20) are

$$w_{ij}^* = \frac{w_i(\hat{F}_i - 1)}{\sum_{i\in s} r_i w_i(\hat{F}_i - 1)}. \tag{21}$$

Using the fractional weights (21) ensures that the imputation variance is eliminated and $\hat{Y}_{\mathrm{MRF}}$ is said to be fully efficient, a term coined by Kim & Fuller (2004). Its asymptotic properties are thus identical to those of (12). As a result, the estimator $\hat{Y}_{\mathrm{MRF}}$ is multiply robust.

7. Simulation study

We performed a simulation study to assess the bias and efficiency of the multiply robust estimator $\hat{Y}_{\mathrm{MR}}$ and the relative bias and coverage probability of the jackknife variance estimator $\hat{V}_{\mathrm{J}}$. We considered the simulation set-up of Kang & Schafer (2007), also used by Chan & Yam (2014) and Han (2014a), among others. We generated $B = 1000$ finite populations of size $N = 10\,000$ as follows. For each unit, a vector $x = (x_1, x_2, x_3, x_4)^{\mathrm{T}}$ was generated from a standard multivariate normal distribution. Then, the characteristic of interest y was generated as $y = 210 + 27{\cdot}4x_1 + 13{\cdot}7(x_2 + x_3 + x_4) + \epsilon$, where the errors ϵ were standard normal. Finally, for each population unit, we generated a size variable $\psi = 0{\cdot}5\chi + 1$, where χ was drawn from a chi-square distribution with one degree of freedom.

From each finite population, we selected a sample of size $n = 800$ using randomized systematic sampling with probability proportional to size. That is, the inclusion probability attached to unit i was defined as $\pi_i = n\psi_i / \sum_{j\in U} \psi_j$.

In each sample, nonresponse to item y was generated with probability $p_i = \{1 + \exp(\alpha_0 + \alpha_1 x_{1i} + \alpha_1 x_{2i} + \alpha_3 x_{3i} + \alpha_4 x_{4i})\}^{-1}$. We used three sets of values for $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4)$: $(-1, 1, -0{\cdot}5, 0{\cdot}25, 0{\cdot}1)$, $(1, -0{\cdot}5, 0{\cdot}25, 0{\cdot}1)$ and $(1{\cdot}1, 1, -0{\cdot}5, 0{\cdot}25, 0{\cdot}1)$. These led to response rates of 30%, 50% and 70%, respectively.

As in Kang & Schafer (2007), we considered the following transformations of the x-variables: $z_1 = \exp(x_1/2)$, $z_2 = x_2\{1 + \exp(x_1)\}^{-1} + 10$, $z_3 = (x_1x_3/25 + 0{\cdot}6)^3$ and $z_4 = (x_2 + x_4 + 20)^2$. The correct nonresponse and imputation models for $p(x)$ and $m(x)$ were estimated by $p^1(x;\alpha^1) = \left\{1 + \exp\left(\alpha_0^1 + \alpha_1^1 x_1 + \alpha_2^1 x_2 + \alpha_3^1 x_3 + \alpha_4^1 x_4\right)\right\}^{-1}$ and $m^1(x;\beta^1) = \beta_0^1 + \beta_1^1 x_1 + \beta_2^1 x_2 + \beta_3^1 x_3 + \beta_5^1 x_4$. The incorrect models were estimated by $p^2(z;\alpha^2)$ and $m^2(z;\beta^2)$, where $p^2(z;\alpha^2)$ and $m^2(z;\beta^2)$ were obtained from $p^1(x;\alpha^1)$ and $m^1(x;\beta^1)$ after replacing x with $z = (z_1, z_2, z_3, z_4)^{\mathrm{T}}$.

We were interested in estimating the finite population mean, $\bar{Y} = N^{-1}\sum_{i\in U} y_i$. Because the proposed imputation procedure may be based on different combinations of the models, we use four digits between parentheses to distinguish estimators constructed using different models. The first two digits correspond to the correct and incorrect nonresponse models, respectively, and the last two digits correspond to the correct and incorrect imputation models, respectively. For example, the estimator $\bar{y}(1010)$ is based on the nonresponse model $p^1(x;\alpha^1)$ and the imputation model $m^1(x;\beta^1)$, while $\bar{y}_{\mathrm{MR}}(1111)$ denotes the estimator based on the four models.

We computed the following estimators of $\bar{Y}$: the complete-data estimator $\bar{y}_{\mathrm{COM}} = \sum_{i\in s} w_i y_i / \sum_{i\in s} w_i$, which assumes no missing values; four doubly robust estimators of the form $\bar{y}_{\mathrm{DR}} = \hat{Y}_{\mathrm{DR}} / \sum_{i\in s} w_i$ where $\hat{Y}_{\mathrm{DR}}$ is described in § 2, namely $\bar{y}_{\mathrm{DR}}(1010)$, $\bar{y}_{\mathrm{DR}}(1001)$, $\bar{y}_{\mathrm{DR}}(0110)$ and $\bar{y}_{\mathrm{DR}}(0101)$; two imputed estimators based on nearest-neighbour imputation (Chen & Shao, 2000), $\bar{y}_{\mathrm{NN}} = \hat{Y}_{\mathrm{I}} / \sum_{i\in s} w_i$ where $\hat{Y}_{\mathrm{I}}$ is given by (1), i.e., $\bar{y}_{\mathrm{NN}}(0010)$ and $\bar{y}_{\mathrm{NN}}(0001)$. The missing value y_i was replaced with the value of the closest respondent value based on Euclidean distance with respect to a set of matching variables. The estimators $\bar{y}_{\mathrm{NN}}(0010)$ and $\bar{y}_{\mathrm{NN}}(0001)$ were based on the set of matching variables x and z, respectively. Nearest-neighbour imputation is robust with respect to the misspecification of the link function of the imputation model. We also considered two imputed estimators based on random hot-deck imputation within classes, $\bar{y}_{\mathrm{HD}} = \hat{Y}_{\mathrm{I}} / \sum_{i\in s} w_i$

Table 1. *Relative bias (%), standard error and root mean squared error of several estimators with different response rates*

Response rate		30%			50%			70%	
Estimator	RB(%)	SE	RMSE	RB(%)	SE	RMSE	RB(%)	SE	RMSE
$\bar{y}_{\text{COM}}$	0·02	1·38	1·38	−0·01	1·40	1·40	−0·02	1·43	1·43
$\bar{y}_{\text{DR}}(1010)$	0·02	1·38	1·38	−0·01	1·40	1·40	−0·02	1·43	1·43
$\bar{y}_{\text{DR}}(1001)$	0·13	2·69	2·70	0·04	1·97	1·97	0·01	1·58	1·58
$\bar{y}_{\text{DR}}(0110)$	0·16	11·47	11·47	0·00	1·45	1·45	−0·02	1·43	1·43
$\bar{y}_{\text{DR}}(0101)$	−23·47	704·55	706·27	−5·01	36·15	37·65	−2·48	16·06	16·88
$\bar{y}_{\text{NN}}(0010)$	−1·26	1·48	3·03	−0·79	1·44	2·19	−0·41	1·44	1·68
$\bar{y}_{\text{NN}}(0001)$	−6·16	2·16	13·11	−3·83	1·70	8·21	−1·96	1·52	4·39
$\bar{y}_{\text{HD}}(1000)$	−0·48	2·25	2·46	−0·44	1·72	1·96	−0·30	1·61	1·73
$\bar{y}_{\text{HD}}(0100)$	−1·42	2·28	3·76	−1·28	1·87	3·27	−0·95	1·66	2·59
$\bar{y}_{\text{MR}}(1010)$	0·02	1·38	1·38	−0·01	1·40	1·40	−0·02	1·43	1·43
$\bar{y}_{\text{MR}}(1001)$	0·14	1·95	1·97	0·06	1·63	1·63	0·02	1·51	1·51
$\bar{y}_{\text{MR}}(0110)$	0·02	1·38	1·38	−0·01	1·40	1·40	−0·02	1·43	1·43
$\bar{y}_{\text{MR}}(0101)$	−1·47	2·05	3·70	−1·20	1·73	3·05	−0·76	1·55	2·22
$\bar{y}_{\text{MR}}(1110)$	0·02	1·38	1·38	−0·01	1·40	1·40	−0·02	1·43	1·43
$\bar{y}_{\text{MR}}(1101)$	0·10	1·95	1·96	0·03	1·64	1·64	−0·01	1·51	1·51
$\bar{y}_{\text{MR}}(1011)$	0·02	1·38	1·38	−0·01	1·40	1·40	−0·02	1·43	1·43
$\bar{y}_{\text{MR}}(0111)$	0·02	1·38	1·38	−0·01	1·40	1·40	−0·02	1·43	1·43
$\bar{y}_{\text{MR}}(1111)$	0·02	1·38	1·38	−0·01	1·40	1·40	−0·02	1·43	1·43

RB, relative bias; SE, standard error; RMSE, root mean squared error.

where $\hat{Y}_{\text{I}}$ is given by (1): $\bar{y}_{\text{HD}}(1000)$ and $\bar{y}_{\text{HD}}(0100)$. Preliminary estimated response probabilities were first obtained using a logistic regression model based on a set of covariates, and then the sample was partitioned into six equal-size classes according to the preliminary probabilities. A missing value within a class was replaced by the value of a respondent selected with replacement within the same class with equal selection probabilities (Little, 1986; Haziza & Beaumont, 2007). The estimators $\bar{y}_{\text{HD}}(1000)$ and $\bar{y}_{\text{HD}}(0100)$ were obtained by fitting a logistic model based on the sets of predictors x and z, respectively. As for nearest-neighbour imputation, this version of random hot-deck imputation within classes is nonparametric and is robust with respect to the misspecification of the link function of the nonresponse model. Finally, we computed nine multiply robust estimators of the form $\bar{y}_{\text{MR}} = \hat{Y}_{\text{MR}} / \sum_{i \in s} w_i$ based on the generalized pseudo-empirical likelihood distance function, where $\hat{Y}_{\text{MR}}$ is given by (12): $\bar{y}_{\text{MR}}(1010), \bar{y}_{\text{MR}}(1001), \bar{y}_{\text{MR}}(0110), \bar{y}_{\text{MR}}(0101), \bar{y}_{\text{MR}}(1110), \bar{y}_{\text{MR}}(1101), \bar{y}_{\text{MR}}(1011), \bar{y}_{\text{MR}}(0111)$ and $\bar{y}_{\text{MR}}(1111)$. The first four estimators were based on a single nonresponse model and a single imputation model, which makes them directly comparable to the doubly robust estimators based on the same models.

We computed the Monte Carlo relative bias, standard error and root mean squared error of each estimator. The results, presented in Table 1, correspond to response rates of 30%, 50%, and 70%.

The complete-data estimator $\bar{y}_{\text{COM}}$ showed negligible bias in all the scenarios, and was more efficient than all the other estimators, which can be explained by the fact that it did not suffer from the additional variability due to nonresponse.

The doubly robust estimators $\bar{y}_{\text{DR}}$ showed negligible bias when either model was correctly specified, consistent with the theory. When both models were misspecified, the estimator $\bar{y}_{\text{DR}}(0101)$ was significantly biased. The magnitude of the bias increased as the nonresponse rate increased. A comparison of $\bar{y}_{\text{DR}}$ and $\bar{y}_{\text{MR}}$ based on a single nonresponse model and a single imputation model

Table 2. *Coverage rate* (%) *and relative bias* (%) *of jackknife variance estimators with a response rate of* 50%

Estimator	$\bar{y}_{\rm MR}\,(1110)$	$\bar{y}_{\rm MR}\,(1101)$	$\bar{y}_{\rm MR}\,(1011)$	$\bar{y}_{\rm MR}\,(0111)$	$\bar{y}_{\rm MR}\,(1111)$
CR	94	95	94	94	94
RB	−6·08	7·63	−6·07	−6·06	−6·02

CR, coverage rate; RB, relative bias.

suggests that $\bar{y}_{\rm MR}$ was more efficient than the corresponding version of $\bar{y}_{\rm DR}$ in most scenarios. For instance, for a response rate of 50%, the root mean squared error of $\bar{y}_{\rm DR}\,(1001)$ was equal to 1·97, whereas it was equal to 1·63 for $\bar{y}_{\rm MR}\,(1001)$. The results were particularly interesting when both models were misspecified. In this case, $\bar{y}_{\rm MR}\,(0101)$ was less biased and was significantly more efficient than $\bar{y}_{\rm DR}\,(0101)$. For a response rate of 50%, the relative bias of $\bar{y}_{\rm DR}\,(0101)$ was approximately equal to −5·01%, whereas $\bar{y}_{\rm MR}\,(0101)$ showed a relative bias of approximately −1·20%. Also, the standard error of $\bar{y}_{\rm MR}\,(0101)$ was considerably smaller than that of $\bar{y}_{\rm DR}\,(0101)$, with values respectively equal to 1·73 and 36·15.

When the underlying models were correctly specified, both random hot-deck imputation within classes and nearest-neighbour imputation led to imputed estimators with lower efficiency than that of the multiply robust estimators for most scenarios. When the models were not correctly specified, both imputation procedures led to biased estimators with large root mean squared error. These results suggest that, although nearest-neighbour imputation and random hot-deck imputation within classes are robust with respect to the misspecification of the link function of the underlying model, they do not have good numerical properties if the set of predictors is misspecified.

The multiply robust estimators based on at least three models showed no bias and were efficient in all the scenarios except one, as $\bar{y}_{\rm MR}\,(1101)$ was somehow less efficient than the other multiply robust estimators. For instance, for a response rate of 50%, the standard error of $\bar{y}_{\rm MR}\,(1101)$ was approximately equal to 1·64, whereas it was approximately equal to 1·39 for the other estimators.

We assessed the performance of the proposed jackknife variance estimator (14) in terms of relative bias and coverage probability of normal confidence intervals. We present the results corresponding to a response rate of 50% with $n = 200$. Table 2 presents the results for a subset of the estimators considered above. Other scenarios led to to very similar results. The jackknife variance estimator performed very well in terms of relative bias in all the scenarios, with an absolute relative bias no larger than 7·63%. The coverage rates were all very close to the nominal 95%. The results suggest that the proposed confidence interval is multiply robust in the sense that the coverage probability is close to the nominal rate if all models but one are misspecified.

Finally, to assess the performance of multiply robust estimators when all the models are misspecified, we ran additional scenarios and computed additional estimators for a nonresponse rate of 50%. The first estimator, denoted by $\bar{y}_{\rm MR1}$, was based on the misspecified models $p^2(z;\alpha^2)$ and $m^2(z;\beta^2)$. In Table 1, this estimator was denoted by $\bar{y}_{\rm MR}\,(0101)$. The second estimator, $\bar{y}_{\rm MR2}$, was based on $p^2(z;\alpha^2)$ and $m^2(z;\beta^2)$ as well as an additional imputation model that included (z_1, z_2, z_3, z_4) and all their interactions as predictors. The third estimator, $\bar{y}_{\rm MR3}$, was based on $p^2(z;\alpha^2)$ and $m^2(z;\beta^2)$ as well as an additional imputation model that included $(z_1^{1/2}, z_2^{1/2}, z_3^{1/2}, z_4^{1/2})$ and all their interactions as predictors. The fourth estimator, $\bar{y}_{\rm MR4}$, was based on $p^2(z;\alpha^2)$ and $m^2(z;\beta^2)$ as well as an additional imputation model that included $\{\log(z_1), \log(z_2), \log(z_3), \log(z_4)\}$ and all their interactions as predictors. The fifth estimator, $\bar{y}_{\rm MR5}$, was based on $p^2(z;\alpha^2)$ and $m^2(z;\beta^2)$ as well as all the models considered in $\bar{y}_{\rm MR2}$ to $\bar{y}_{\rm MR4}$.

Table 3. *Relative bias* (%), *standard error and root mean squared error of additional multiply robust estimators with a response rate of* 50%

Estimator	$\bar{y}_{\text{MR1}}$	$\bar{y}_{\text{MR2}}$	$\bar{y}_{\text{MR3}}$	$\bar{y}_{\text{MR4}}$	$\bar{y}_{\text{MR5}}$
RB	−1·20	−0·49	−0·52	−0·55	−0·42
SE	1·73	1·75	1·57	1·51	1·70
RMSE	3·05	2·03	1·91	1·90	1·92

For $\bar{y}_{\text{MR3}}$ and $\bar{y}_{\text{MR4}}$, transforming the predictors was justified by the fact that the variables z_1, z_2, z_3 and z_4 exhibited skewed distributions; see also Chan & Yam (2014). The results are presented in Table 3.

The estimators $\bar{y}_{\text{MR2}}$ to $\bar{y}_{\text{MR5}}$ showed less bias and a lower root mean squared error than that of $\bar{y}_{\text{MR1}}$. Including models with transformed predictors led to lower bias and increased efficiency. The estimators $\bar{y}_{\text{MR3}}$ to $\bar{y}_{\text{MR5}}$ were much more efficient than $\bar{y}_{\text{MR1}}$, with a root mean squared error approximately equal to 1·90 for $\bar{y}_{\text{MR3}}$ to $\bar{y}_{\text{MR5}}$ compared with 3·05 for $\bar{y}_{\text{MR1}}$. These results suggest that multiply robust procedures tend to have very good numerical properties when all the models are misspecified.

Acknowledgement

The work of the second author was supported by the Natural Sciences and Engineering Research Council of Canada and the Canadian Statistical Sciences Institute. The authors wish to thank the editor, an associate editor and two referees for their constructive comments.

Appendix

Proof of Theorem 1

Assume $m(x_i;\beta) = a_0 + \sum_{k=1}^{K} a_k m^k(x_i;\beta^k)$. Then, under regularity conditions, it can be shown that $\hat{\gamma}_p = \gamma_p^* + o_{\text{p}}(1)$, where $\gamma_p^* = (a_0^*, 0, \ldots, 0, a_1, a_2, \ldots, a_K)$, with $a_0^* = a_0 + \sum_{k=1}^{K} a_k E\left\{m^k(x;\beta^k)\right\}$. Therefore,

$$
\begin{aligned}
\frac{\hat{Y}_{\text{MR}}}{Y} &= \frac{N^{-1}\left\{\sum_{i\in s} r_i w_i y_i + \sum_{i\in s}(1-r_i)\, w_i h_i^{\mathrm{T}}\hat{\gamma}_p\right\}}{\bar{y}} \\
&= \frac{E\left[N^{-1}\left\{\sum_{i\in s} r_i w_i y_i + \sum_{i\in s}(1-r_i)\, w_i m(x_i;\beta)\right\}\right]}{E\{\bar{y}\}} + o_{\text{p}}(1) \\
&= \frac{E\{r_i y_i + (1-r_i) m(x_i;\beta)\}}{E(Y)} + o_{\text{p}}(1) \\
&= 1 + o_{\text{p}}(1), \qquad n, N \to \infty.
\end{aligned}
$$

Proof of Theorem 2

Without loss of generality, we assume that the model $p^1(x_i;\alpha^1)$ is correctly specified. Because the solution to minimizing the distance function subject to (5)–(7) is unique, under regularity conditions, we have $\hat{\gamma}_p = \gamma_p^* + o_{\mathrm{p}}(1)$, where $\lambda_r^* = (\lambda_0^*, 1, 0, \ldots, 0)$ with $\lambda_0^* = E\left[L\left\{1/p^1(x_i;\alpha^1)\right\}\right]$. It follows that

$$\begin{aligned} F(\hat{\lambda}_r^{\mathrm{T}} h_i) &= F\left\{\hat{\lambda}_0 + \hat{\lambda}_1(\hat{L}_i^1 - \hat{L}^1) + \cdots + \hat{\lambda}_{J+K}(m_i^K - \hat{m}^K)\right\} \\ &= F\left\{F^{-1}(1/p_i^1)\right\} + o_{\mathrm{p}}(1) \\ &= 1/p_i^1 + o_{\mathrm{p}}(1), \qquad n, N \to \infty. \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{\hat{Y}_{\mathrm{MR}}}{Y} &= \frac{N^{-1}\left\{\sum_{i\in s_r} w_i \hat{F}_i y_i + \left(\sum_{i\in s} w_i h_i - \sum_{i\in s_r} w_i \hat{F}_i h_i\right)^{\mathrm{T}} \hat{\gamma}_p\right\}}{\bar{y}} \\ &= \frac{E\left[N^{-1}\left\{\sum_{i\in s_r} w_i/p_i^1 y_i + \left(\sum_{i\in s} w_i h_i - \sum_{i\in s_r} w_i/p_i^1 h_i\right)^{\mathrm{T}} \gamma_p^*\right\}\right]}{E(Y)} + o_{\mathrm{p}}(1) \\ &= 1 + o_{\mathrm{p}}(1), \qquad n, N \to \infty. \end{aligned}$$

Proof of Theorem 3

We assume the following regularity conditions:

Condition A1. $(\theta^*, \lambda^*, \gamma^*)$ are the unique probability limits of $(\hat{\theta}, \hat{\lambda}_r, \hat{\gamma}_p)$ and $\hat{\theta}$ has influence function

$$\hat{\theta} - \theta^* = \frac{1}{N}\sum_{i\in s} w_i Q(x_i;\theta^*) + o_{\mathrm{p}}(n^{-1/2}),$$

where $\theta^* = (\alpha^{1*}, \ldots, \alpha^{J*}, \beta^{1*}, \ldots, \beta^{K*})$, $\hat{\theta} = (\hat{\alpha}^1, \ldots, \hat{\alpha}^J, \hat{\beta}^1, \ldots, \hat{\beta}^K)$ and

$$Q(x_i;\theta^*) = -\left[E\left\{N^{-1}\frac{\partial S(\theta^*)}{\partial\theta}\right\}\right]^{-1}\left(s_{1i}^1, \ldots, s_{1i}^J, s_{2i}^1, \ldots, s_{2i}^K\right)^{\mathrm{T}},$$

with $S(\theta) = \left\{S_1^1(\alpha^1), \ldots, S_J^1(\alpha^J), S_2^1(\beta^1), \ldots, S_2^K(\beta^K)\right\}^{\mathrm{T}}$ and

$$s_{1i}^j = \frac{r_i - p^j(x_i;\alpha^j)}{p^j(x_i;\alpha^j)\left\{1 - p^j(x_i;\alpha^j)\right\}}\frac{\partial p^j(x_i;\alpha^j)}{\partial\alpha^j}, \quad s_{2i}^k = r_i\left\{y_i - m^k(x_i;\beta^k)\right\}\frac{\partial m^k(x_i;\beta^k)}{\partial\beta};$$

Condition A2. $F(t)$ has continuous first derivative and $h(x;\theta)$ has continuous first derivative with respect to θ;

Condition A3. $E(Y^2), E(F^2), E(h^2), E(S^2), E(\dot{F})$ and $E(\dot{h})$ are bounded;

Condition A4. The true response probability satisfies $0 < a < p(x;\alpha) < 1$ almost surely.

Define

$$U(\hat{\lambda}_r) = \sum_{i\in s_r} w_i F(\hat{\lambda}_r^{\mathrm{T}} h_i) h_i - \sum_{i\in s} w_i h_i.$$

Then, by using Taylor linearization and Conditions A1 and A2, it can be shown that

$$\hat{\lambda}_r^{\mathrm{T}} - \lambda^* = -\left\{E\left(\frac{\partial U}{\partial\hat{\lambda}_r}\right)\right\}^{-1} U(\lambda^*) + o_{\mathrm{p}}(n^{-1/2}). \qquad \text{(A1)}$$

By Taylor linearization, Condition A1 and (A1) and because $\sum_{i\in s} w_i \left(r_i\hat{F}_i - 1\right) h_i^{\mathrm{T}} = 0$, we have

$$\begin{aligned}
\hat{Y}_{\mathrm{MR}}\left(\hat{\theta}, \hat{\lambda}_r\right) &= \hat{Y}_\pi + \sum_{i\in s} w_i \left(r_i\hat{F}_i - 1\right) h_i^{\mathrm{T}} \left(\gamma^* - \hat{\gamma}_p\right) + \sum_{i\in s} w_i \left(r_i\hat{F}_i - 1\right) \hat{e}_i \\
&= \hat{Y}_\pi + \sum_{i\in s} w_i \left(r_i\hat{F}_i - 1\right) \hat{e}_i \\
&= \hat{Y}_\pi + \sum_{i\in s} w_i \left(r_i F_i - 1\right) e_{i,0} + \sum_{i\in s} w_i r_i \dot{F}_i h_{i,0} e_{i,0} \left(\hat{\lambda}_r - \lambda^*\right) \\
&\quad + \left\{\sum_{i\in s} w_i r_i \dot{F}_i \lambda^* \dot{h}_{i,0} e_{i,0} - \sum_{i\in s} w_i \left(r_i F_i - 1\right) \gamma^* \dot{h}_{i,0}\right\} \left(\hat{\theta} - \theta^*\right) + o_{\mathrm{p}}\left(n^{-1}N\right) \\
&= \sum_{i\in s} w_i \eta_{i,0} + o_{\mathrm{p}}\left(n^{-1}N\right),
\end{aligned}$$

where $\hat{e}_i = y_i - h_i^{\mathrm{T}}\hat{\gamma}_p$, $\hat{Y}_\pi = \sum_{i\in s} w_i y_i$ and $\eta_{i,0}$ is defined in Theorem 3.

REFERENCES

BANG, H. & ROBINS, J. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61**, 962–73.

BERGER, Y. G. (2007). A jackknife variance estimator for unistage stratified samples with unequal probabilities. *Biometrika* **94**, 953–64.

BERGER, Y. G. (2011). Asymptotic consistency under large entropy sampling designs with unequal probabilities. *Pak. J. Statist.* **27**, 407–26.

CAO, W., TSIATIS, A. A. & DAVIDIAN, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika* **96**, 723–34.

CHAN, K. C. G. & YAM, S. C. P. (2014). Oracle, multiple robust and multipurpose calibration in a missing response problem. *Statist. Sci.* **29**, 380–6.

CHEN, H. L. & SHAO, J. (2000). Nearest-neighbour imputation for survey data. *J. Offic. Statist.* **16**, 583–99.

DEVILLE, J. C. & SÄRNDAL, C. E. (1992). Calibration estimators in survey sampling. *J. Am. Statist. Assoc.* **87**, 376–82.

FAY, R. E. (1991). A design-based perspective on missing data variance. *Proc. 1991 Annual Research Conf.* Washington, DC: US Bureau of the Census, pp. 429–40.

FOX, J. (2008). *Applied Regression Analysis and Generalized Linear Models*. Los Angeles: Sage, 2nd ed.

FULLER, W. A. & KIM, J. K. (2005). Hot deck imputation for the response model. *Survey Methodol.* **31**, 139–49.

HAN, P. (2014a). A further study of the multiply robust estimator in missing data analysis. *J. Statist. Plan. Infer.* **148**, 101–10.

HAN, P. (2014b). Multiply robust estimation in regression analysis with missing data. *J. Am. Statist. Assoc.* **109**, 1159–73.

HAN, P. & WANG, L. (2013). Estimation with missing data: Beyond double robustness. *Biometrika* **100**, 417–30.

HAZIZA, D. (2009). Imputation and inference in the presence of missing data. *In Handbook of Statistics, Volume 29A, Sample Surveys: Theory Methods and Inference,* D. Pfeffermann and C. R. Rao, eds. Amsterdam: Elsevier, pp. 215–46.

HAZIZA, D. & BEAUMONT, J. F. (2007). On the construction of imputation classes in surveys. *Int. Statist. Rev.* **75**, 25–43.

HAZIZA, D. & RAO, J. N. K. (2006). A nonresponse model approach to inference under imputation for missing survey data. *Survey Methodol.* **32**, 53–64.

HAZIZA, D., NAMBEU, C. O. & CHAUVET, G. (2014). Doubly robust imputation procedures for finite population means in the presence of a large number of zeroes. *Can. J. Statist.* **42**, 650–69.

HE, Y. & RAGHUNATHAN, T. E. (2006). Tukey's *gh* distribution for multiple imputation. *Am. Statist.* **60**, 251–6.

ISAKI, C. T. & FULLER, W. A. (1982). Survey design under the regression superpopulation model. *J. Am. Statist. Assoc.* **77**, 89–96.

KANG, J. D. Y. & SCHAFER, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data (with discussion). *Statist. Sci.* **22**, 523–39.

KIM, J. K. & FULLER, W. A. (2004). Fractional hot-deck imputation. *Biometrika* **91**, 559–78.

KIM, J. K. & HAZIZA, D. (2014). Doubly robust inference with missing data in survey sampling. *Statist. Sinica* **24**, 375–94.

KIM, J. K. & PARK, H. A. (2006). Imputation using response probability. *Can. J. Statist.* **34**, 171–82.

Kim, J. K. & Rao, J. N. K. (2009). Unified approach to linearization variance estimation from survey data after imputation for item nonresponse. *Biometrika* **96**, 917–32.
Kott, P. S. (1994). A note on handling nonresponse in sample surveys. *J. Am. Statist. Assoc.* **89**, 693–6.
Kott, P. S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodol.* **32**, 133–42.
Kott, P. S. & Chang, T. (2010). Using calibration weighting to adjust for nonignorable unit nonresponse. *J. Am. Statist. Assoc.* **105**, 1265–75.
Léger, C., Politis, D. N. & Romano, J. P. (1992). Bootstrap technology and applications. *Technometrics* **34**, 378–98.
Little, R. J. A. (1986). Survey nonresponse adjustments. *Int. Statist. Rev.* **54**, 139–57.
Rao, J. N. K. (1965). On two simple schemes of unequal probability sampling without replacement. *J. Indian Statist. Assoc.* **3**, 173–80.
Robbins, M. W., Sujit, K. & Habiger, J. D. (2013). Imputation in high-dimensional economic data as applied to the Agricultural Resource Management Survey. *J. Am. Statist. Assoc.* **108**, 81–95.
Robins, J. M., Rotnitzky, A. & Zhao, L. P. (1994). Estimation of regression coefficient when some regressors are not always observed. *J. Am. Statist. Assoc.* **89**, 846–66.
Robins, J. M., Sued, M., Gomez-Lei, Q. & Rotnitzky, A. (2007). Comment: Performance of double-robust estimators when inverse probability weights are highly variable. *Statist. Sci.* **22**, 544–59.
Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–92.
Sampford, M. R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika* **54**, 499–513.
Scharfstein, D. O., Rotnitzky, A. & Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion and rejoinder). *J. Am. Statist. Assoc.* **94**, 1096–146.
Shao, J. & Steel, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *J. Am. Statist. Assoc.* **94**, 254–65.
Tan, Z. (2006). A distributional approach for causal inference using propensity scores. *J. Am. Statist. Assoc.* **101**, 1619–37.
Tillé, Y. (2006). *Sampling Algorithms*. New York: Springer.
Wand, M. P. & Jones, M. C. (1995). *Kernel Smoothing*. London: Chapman and Hall.
Weisberg, S. (1985). *Applied Linear Regression*. New York: Wiley, 1st ed.
Wu, C. & Sitter, R. R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *J. Am. Statist. Assoc.* **96**, 185–93.
Wu, C. & Lu, W. W. (2016). Calibration weighting methods for complex surveys. *Int. Statist. Rev.* **84**, 79–98.

[*Received on* 21 *May* 2015. *Editorial decision on* 2 *January* 2017]